DOCUMENT RESUME

ED 047 546 FL 001 498

AUTHOR Hanna, Ian
TITLE Language Aptitude Testing and Streaming.
INSTITUTION Australian Association of Adult Education, Melbourne.
PUB DATE Nov 68
NOTE 12p.; Reprint from the Australian Journal of Adult Education, v8 n3 p122-132 Nov 1968

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Achievement Rating, *Adult Education, *Aptitude Tests, Cognitive Ability, Educational Background, Homogeneous Grouping, *Language Instruction, *Language Tests, Standardized Tests, *Student Evaluation, Test Validity
IDENTIFIERS Australia, *Modern Language Aptitude Test

ABSTRACT

Two methods used in the determination of student aptitude in adult language instruction programs in Australia are described and ccmpared in this article. The first method screens potential students into homogeneous classes on the basis of previous education, while the second utilizes test results derived from the administration of the short form of the Modern Language Aptitude Test. Test results are correlated with language tutors' ratings of students. Procedures and results of the screening processes are discussed and summarized in several tables. (NL)

EDO 47546

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Australian Journal of Adult Education TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

LANGUAGE APTITUDE TESTING AND STREAMING

by
Ian Hanna

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

FL 001 498

Reprinted from AUSTRALIAN JOURNAL OF ADULT EDUCATION, VOL. VIII, No. 3, November, 1968, pages 122-132.

LANGUAGE APTITUDE TESTING AND STREAMING

Ian Hanna

Foreign language teaching has, since the war, become one of the biggest parts of Australian adult education. It is provided through all major forms of adult education—universities, statutory authorities, state education departments and voluntary bodies. As the enrolments are a very substantial part of adult education classes, it has been all too evident that language classes have a very big drop-out. A fall from first- to second-year enrolments of 50% is quite normal, though this varies appreciably for different institutions and subjects.[1]

In adult education language classes the most important factors contributing to the students' success and satisfaction, are (a) the tutors' competence in working with adults, (b) the appropriateness of the courses to adult needs and interests. In the last decade, in Australia, the only major widespread, improvement in adult education language facilities has been the part-time use of the language laboratories belonging either to universities or other institutions of a similar kind. The more fundamental improvements, the design of courses and the training of adult education tutors, has been, in the main, beyond the resources available to non-vocational adult education institutions. The organisers and tutors of adult education language classes must make such improvements as they can with the limited resources provided for their use. At the same time, it may be possible to reduce the drop-out figures if those in charge can stream the students, to produce classes which are more suitable to the students' motives, aptitudes, and linguistic background.

Though, in most institutions, a majority of language students have had previous language experience in their secondary education,[2] it is often many years since they left school. Their schools, as a rule, gave a limited experience of language skills in the form of contrived translation exercises and grammatical rule learning. But they also provided better opportunities for learning, one period a day and compulsory homework. It is understandable that under these conditions, a number of students achieved a reasonable success in language exams.

A most experienced adult educationist, H. A. Jones,[3] formerly of the City Literary Institute, London (which maintains one of the largest and best-organised adult education language schools outside the universities), has expressed the opinion that many adults enrol in languages because they wish to recapture the sense of early and comparatively easy success of their schooldays. Under typical adult education conditions these adults receive only one session per week after their day's work, and they are required to do considerable individual study in their own largely non-existent spare time. While the school student may have little interest in the language he is studying successfully, he has usually

an enforced attendance, and an incentive which is not primarily the interest in the language, but merely the making of a score in an examination competition. The adult student, however, has to provide all his own objectives, though these may sometimes be supplemented by the incentive of professional advancement or the need to speak to New Australian in-laws. If the adult is not satisfied, he soon remembers that there are many other easier ways of spending his time, and he leaves the class.

Adult education classes are thus highly sensitive barometers of the despair and dissatisfaction of their students, whereas a school class, once assembled, is hardly sensitive at all. However, since languages have more and more ceased to be compulsory in secondary schools, the school students have been increasingly exercising their option to drop these subjects altogether. This has been of concern to the language teaching profession for some time, and there does not appear to be any sign that the drift from secondary school languages is likely to stop.[4]

The teaching of adult languages differs substantially from the tutoring of traditional liberal adult education subjects. The essential difference is that for a considerable period of elementary study, which may be a matter of years, the student must set about mastering a large body of knowledge and acquiring new skills. In this period there is comparatively little contribution that the individual can make to the class by way of discussion and by the sharing of experience and opinions which help to create personal interest in the liberal tutorial.[5] For this reason, elementary language classes have somewhat less chance of sustaining interest than most other adult education studies, unless it is the interest in achievement reinforced by success.

It is generally agreed that the teaching of a substantial body of knowledge and definite skills can be better carried out in classes that are more or less homogeneous, so that the tutor can assume that he knows the general level of skill and understanding of his pupils, and can plan a lesson which will add the next acceptable stages to this knowledge and skill. Ideally the full-time secondary school is supposed to provide this condition, and largely does, as the teacher can assume that those who enter a class have passed the previous year's work, and gained at least competence up to an identifiable level.

First year adult language classes are seldom homogeneous in matters of language or scholastic aptitude, previous level of languages learnt, motives or sophistication. There is little experimental data to verify the assumption that homogeneity in our adult education language classes improves performance and re-enrolment because there has not been a substantial enough body of students in any one institution on which to conduct a well-controlled experiment. But in spite of this absence of exact data, we are still probably justified in assuming, *pro tem*, that homogeneous classes would be better.

In some cases it has been possible to make a language class intake more consistent by such devices as stating desired pre-

requisites of previous language standards or clearly stating the objectives to be reached by a course, i.e. Russian for Scientists, Japanese for Businessmen, Conversational French, but in most cases this has not been possible. The practical difficulty is, to start with, that a potential enrolment in a language must be rather large by our standards to allow the enrolment to be split into classes of different levels, or precise objectives. In some of the earlier years of the most popular languages, however, this may be tried.

Screening and Streaming

There are two probable conditions in which it might be useful to screen out or stream students. One is where a class can only take a limited enrolment, or where the objective is to produce the highest quality of skill, regardless of the numbers which can be taught. This is more likely to be the case of specialised language teaching schools at a high level, such as in the armed forces. In adult education, the restricting of a class may be needed where a language laboratory is involved, and it is desirable not to allocate seats to students who will probably not succeed. The task here is screening to select the best and eliminate the others.

In adult education, the more probable situation would be that of having a large enough enrolment to allow the streaming of students to make several more homogeneous classes, so that the tutor could get a clearer idea of the general level of the student in his class, and adjust curriculum and speed to suit. Even if parallel classes finish their year with different levels of attainment, it is a reasonable assumption that they may have given more value and satisfaction to a greater number of students, which may be shown by a better retention and re-enrolment figure.[6]

In order to attempt streaming, it is desirable to know if it is possible to make any effective assessment of the students' potential before the class starts. We would, therefore, be interested in finding any practical method which it would be comparatively simple to apply under the normal rushed and overworked conditions of adult education establishments with the general multi-purpose, multi-level intake of students.

Two approaches have been made towards learning more about, and improving the effectiveness of the adult education language classes at The University of Adelaide:

(a) Streaming, according to the educational history of the students, in those first-year classes which have had a large enough intake. This has been confined so far to Italian I and German I. The tutors involved have expressed some satisfaction with the results, and a more objective evaluation is being attempted.

(b) A first experiment at testing language learning aptitude. This is described here at greater length for colleagues who may be interested in applying it to their own classes.

Language Aptitude Testing

This experiment was limited to trying to validate on our adult education language population one published test, as a discriminator between students of greater and lesser aptitude at languages, in a general enrolment of adult education students. The Modern Language Aptitude Test was used.

The assumption in the use of the term "Language Aptitude Test" is that it does not test the amount of language which a person has already learned, as an examination purports to do, but that it measures the ability or potential of the person to learn a language. There are some fundamental questions as to the very nature and the number of aptitudes which may be involved in learning different language skills and different types of languages, but they are outside the scope of this experiment and this article. The experiment was a pragmatic attempt to find if the use of a particular test would improve the selection and streaming of students under existing conditions.

The Modern Language Aptitude Test (M.L.A.T.) was developed at Harvard University between 1955 and 1958 by John B. Carrol, a distinguished American expert in linguistics, and Stanley M. Sapon. The M.L.A.T. which has received quite respectful reviews[7] in the U.S.A. was "designed chiefly to provide an indication of an individual's probable degree of success in learning a foreign language. It is particularly useful in predicting success in learning to speak and understand a foreign language, but it is also useful in predicting success in learning to read, write, and translate a foreign language. It is applicable in connection with both modern spoken languages and ancient languages such as Latin or Greek."[8]

The developmental stages of the M.L.A.T. involved its administration to approximately 5000 individuals in the U.S.A. A student's score on the version of the M.L.A.T. used in this experiment could be compared with (a) a percentile table based on 1600 U.S.A. University and servicemen candidates for language training, or (b) a percentile table based on a population of 2500 Australian adult candidates for language training, and high school students from Intermediate, Leaving and Matriculation classes; this is the table referred to in this experiment. (A percentile table divides the actual scores of the persons tested into 100 groups of equal sizes from lowest to highest scores. The 50th percentile, called the median, divides the population of scores into the higher half and the lower half.)

The complete M.L.A.T. is recorded on magnetic tape. Each person being tested requires a copy of the reusable Test Booklet and the Practice Answer Sheet; he writes his answers on a special, printed form. The Complete Test has five Parts and takes approximately 1½ hours to administer, but a short version of the M.L.A.T. may be used instead. This consists of Parts III, IV and V and takes approximately half the time. The Short Test has the advantage of not requiring a tape recorder for its admin-

istration, the tester reads aloud the instructions, and times each of the three Parts.

A brief summary of the five Parts of the M.L.A.T. with the designers' opinion of the function of each Part:

Part I. *Number Learning*. Score, 43.

This measures one aspect of the memory component of foreign language aptitude but it could also be strongly influenced by "auditory alertness" and thus be a measure of the auditory comprehension of the person.

Part II. *Phonetic Script*. Score, 30.

This measures the ability to form sound-symbol association, the ability to learn correspondences between speech sounds and orthographic symbols. It is also influenced by the ability to mimic speech sounds.

Part III. *Spelling Clues*. Score, 50.

This measures to some extent existing English vocabulary knowledge and also the ability to form sound-symbol association as in Part II. It must be noted that this Part III is a speed test.

Part IV. *Words in Sentences*. Score, 45.

This measures sensitivity to grammatical structure. It may be influenced by the degree of the person's formal training in grammar, but no grammatical terminology is used.

Part V. *Paired Associates*. Score, 24.

This measures the rote memory aspect of learning a foreign language.

Complete Test, total score = 192.
Short Test, total score = 119.

In the third term of 1966 the Complete Test was administered to a number of classes. The test papers were processed at the R.A.A.F. School of Languages at Point Cook, Victoria, through the courtesy of Wing Commander A. J. Garrick who is now Director of the Institute of Languages, University of N.S.W. Point Cook provided us with the student identification, test scores and percentile rating, but other material, including student's age, was not processed.

The test was completed by 131 students who could be identified by name and checked on the class rolls. Two categories of students had to be eliminated from this sample, (a) the 8 foreign-born and (b) the 12 elderly.

(a) *The Foreign-Born*

As the test had been validated only for people who had English as a native language or who spoke it with near-native fluency, the foreign-born students were identified by the tutors and by their entry in class applications. All foreign-born students were eliminated with the exception of two who were language teachers and had an excellent command of English.

(b) *The Elderly*

When the tutors were shown the students' test scores, one pointed out that most of the elderly, who were amongst her best students, had scored very badly. It was evident that this was mainly because the scoring of the test made no allowance for

increasing age which reduces the ability to handle any situation demanding speed, such as a speed test. As no adjustment could be made in the scoring for age all the students whom the tutors considered to be elderly were eliminated from the sample.

After these eliminations, the samples which remained were:

Sample 1. All students who could be identified; total, 111.[9]

Sample 2. Only those students of sample 1 who could be rated by their tutors, on a three-point scale; total, 92.

It is useful to note that the median scores of each of these two samples was 107, which is very close to the median score of 108 for the 2500 Australian language students and adult candidates, referred to previously.

The Criteria and the Experiments

The aim was to see if this test would discriminate between the more competent and the less competent of our adult education students. Therefore we set up two criteria by which this could be judged in this current student body and we examined the M.L.A.T. results to see if they correlated to the criteria. The two criteria were:

A. Elementary—Advanced Study, and

B. The Tutors' Rating of students for their competence, on a three-point scale.

A. ELEMENTARY—ADVANCED STUDY

Sample 1 was used. Class rolls of 1966, 1967 and 1968 were examined to see which students tested in 1966 did not proceed past first-year language classes. The M.L.A.T. scores of this Elementary group were compared with the M.L.A.T. scores of the Advanced group who went on to second or higher years. Our research hypothesis, i.e. our assumption, was that those with lower aptitude scores would be less likely to get past first year. For purposes of the statistical method we set up and challenged the null hypothesis that there was no significant difference between the M.L.A.T. scores of, (i) those 1966 students who in 1966-68 inclusive studied only at first-year level and, (ii) those who in that period studied at a higher level. Table 1 shows (a) the level of study of the students tested in 1966 and (b) the level they had reached by 1968.

TABLE 1

ELEMENTARY-ADVANCED STUDY

	Level of Study Reached			
	Year I, Only	Year II and Higher		
	Number of Students			
(a) In 1966	76	35	=	111
(b) By 1967 and 1968	41	70	=	111

Using the Median Score of 107 as dividing point produced the following distribution:

M.I.L.A.T. Score	Level of Study Reached Year I, Only	Year II and Higher		
	Number of Students			
Above Median	20	35	=	55
Below Median	21	35	=	56
	41	70		111

Findings: No significant difference appears in the scores of those who persisted to higher levels of study and those who did not go past first year. Therefore we cannot reject the null hypothesis and we must reject our original assumption that the Advanced students would score better than the ones who did not get past elementary classes.

It is, of course, quite possible that the students of higher aptitude had reasons for dropping out which differ from those of the lower aptitude students. For example, the higher aptitude group may have dropped out because they were professional people and were too busy, or they went abroad; the lower aptitude group because they became discouraged by their own incompetence. However, we have not sufficient evidence on this; we only know that the test did not discriminate between the tested Elementary and Advanced students between 1966-68.

B. THE TUTORS' RATING OF STUDENTS

Sample 2. Only the 92 students whom the tutors remembered well enough to rate.

The tutors were not shown the students' scores and were asked to rate the students on a three-point scale according to the competence they had shown in class. The tutors were asked to make their judgments according to the standards they had come to expect in adult education classes.

Rating 1. Very Good Students.

Rating 2. Competent to learn languages as taught in adult education classes.

Rating 3. Not Competent Students.

It would have been better to have made the ratings on a four-point scale for in this instance, as is generally the case, the tutors tended to bunch their students into the centre category. Rating 2. Competent. The four-point scale is usually advisable as it forces the spreading of the judgments; it also makes it easier for the experimenter to reduce the distribution to a two-category scale with near equal numbers in each category. This allows the reduction of the data to a 2 x 2 table which is very handy, especially when dealing with small samples.

The students' scores on both the Complete Test and the Short Test (Parts III, IV and V) were compared with the Tutors' Ratings. The research hypothesis was that students given a high

rating by the tutors would have scored better than those with a low rating. The null hypothesis was that there would be no significant difference between the scores of the students to whom the tutors had given *Rating 1. Very Good,* and the scores of those given *Rating 3. Not Competent.*

Table 2 shows the score distribution using the Complete Test.

TABLE 2
TUTORS' RATINGS AND M.L.A.T. SCORES
Complete Test. Median, 107.

M.L.A.T. Score	Tutors' Rating 3	2	1		
	Number of Students				
Above Median	5	18	23	=	46
Below Median	18	19	9	=	46
					92

Table 2 shows the tendency for the higher rated students to have better M.L.A.T. scores than the lower rated students. A Chi-squared test shows that this difference is significant.

Chi-squared (X^2) = 12.4, Two degrees of freedom.
Probability = <.01.

The same students' scores on the Short Test show a similar tendency.

TABLE 3
TUTORS' RATINGS AND M.L.A.T. SCORES
Short Test. Median, 51.

M.L.A.T. Score	Tutors' Rating 3	2	1		
	Number of Students				
Above Median	6	17	23	=	46
Below Median	17	20	9	=	46
					92

Chi-squared (X^2) = 12.1, Two degrees of freedom.
Probability = <.01.

A probability of less than .01 (1%), means that the differences shown in Tables 2 and 3 could have occurred by chance less than once in a 100 samples. The figures can therefore be taken to represent a genuine difference between the higher scores of the higher rated students and the lower scores of the lower rated students.

Findings: We can accept our research hypothesis and reject the null hypothesis, as the M.L.A.T. scores (both Complete and Short Tests) show significant agreement with the Tutors' Ratings, the Short Test giving practically the same results as the Complete Test.

Evaluation

The Chi-squared results are probably very good for such a sample. Because the tests were given late in the third term we may reasonably assume that the most incompetent of the first-year students had already dropped out—the common wasteful process of self-screening. A population screened this way could be expected to show a less clear-cut distinction than the un-screened population at the time of enrolment.

These experiments were in the nature of a validation of the M.L.A.T. for our adult education population, but it must not be assumed that it is a valid test for all Australian adult education students for some years to come. That can only be established by more widespread testings from time to time.

It was known that the test was not valid for the foreign-born unless they were very fluent in English. It now appears that it is not a valid test of the language aptitude of old students (unless a correction can be made for increasing age), and probably not for anyone who is subject to "tape recorder trauma", that is, anyone who gets flustered by the language laboratory situation and, very likely, anyone who is subject to examination panic. But this very difficulty, once understood, adds another use to the test.

Application

The M.L.A.T. would appear to be of use in an Australian adult education institution if there are enough enrolments for a given class to allow (a) streaming of students into classes of different learning intensity or speed, or, (b) the rejection of the less suitable applicants. This latter condition might apply especially for the elimination of excess applicants for a language laboratory class. In that case it might well be a good test for the elderly as we would want to know their ability to stand up to a machine-paced learning situation. This was the suggestion of one of our tutors who pointed out that most of the elderly scored very badly on the M.L.A.T. even though they were generally amongst the most persistent and successful students in the more traditional classes. But it would be unwise to eliminate students from the language laboratory on the basis of their advanced age alone, as some should be quite capable of benefiting from it.

The Short Test: We are looking for a practical way of improving selection or streaming under normal adult education conditions. The Short Test appears to have only a slightly lower "predictive" value than the Complete Test and it is a more convenient test. It does not require a tape recorder and it can be administered in half the time, an important consideration if it is to be applied to a large, newly assembled first-year intake of students.

Until we have more test results from our Australian adult education students, it would be of greater long-range value if the Complete Test could be widely used. A greater body of M.L.A.T. results together with other criteria of students' performances and the use of more sophisticated statistical analysis would show

whether any parts of the test were better predictors of student performance in the different types of adult classes offering.

Improving Prediction: Of course better predictive value could certainly be obtained by using an appropriate battery of tests and relating results to the educational history and other personal data of the students, but this would seldom be practical under usual adult education conditions.

Prediction from Previous Attainment: The best rule of thumb for the prediction of a student's ability to learn a new language would be his previous history of learning languages, including his own native language. Some defects of using this rule of thumb as the only basis for selection are evident, e.g. if the student's previous language achievement was many years ago, it might be little indication of his present capacity. The language skills he acquired at school may differ basically from those required to master his adult studies. If a student has had little opportunity for education, a latent aptitude is not likely to be shown by easily identified attainments.

However, if a student's educational history is considered in conjunction with his M.L.A.T. results, even without refined statistical treatment, the evaluation should be more effective than if either form of evidence were used alone.

A high score on the M.L.A.T. could be taken as a positive indication of ability to tackle most language studies. A low score should not in itself be taken as sufficient evidence of little aptitude for languages, though it would be grounds for considering the student to be a doubtful prospect for a language laboratory course. A recent, or long, history of competence in the study of languages, including English, would be a good indicator of future success in classes which used similar methods. A low M.L.A.T. score, together with a history of little previous language study, would class an applicant as a poor prospect.

In most cases, one good positive indication in either the M.L.A.T. score or educational history would be sufficient to have an applicant classed as a good prospect. At these extremes the judgment is relatively easy. It is in the middle range of doubtful cases that the adult educator will have to make more arbitrary decisions.

It is not enough for these methods of selection to produce more effective classes; it is necessary, also, to have records of the information about the students on which the selection decisions are made. Only in this way can we ensure that the improvements will be continuous and that experiences in one institution may be of help to another one. In this testing situation, as in any other, the validity of the method cannot be better than the reliability.

Therefore if other Australian adult education bodies think they can improve their classes by these suggested methods, or any others, it would be valuable to all concerned to establish a systematic exchange of detailed information on the experiments planned or completed—including the failures.

References and Notes

[1] A 50% drop-out is a conservative estimate but figures are hard to compare because of the different bases of calculation in different institutions.
RIVERS, W., in "Methodology to Achieve the Objectives of Adult Classes", in *Language Teaching for Adult Education,* p. 2, A.A.A.E., Melbourne, 1965, gives a generalisation "above 50%". Rivers's articles in this volume are very good contributions by an experienced teacher of adult education languages.
LAWRENCE, F.: "A Student Survey", *A.J.A.E.*, Vol. VII, No. 1, July, 1967, p. 43, gives "50% plus" as normal for first year but her criterion of drop-out is more stringent than some others.

[2] BODI, L.: "Some Problems of Language Teaching in Australian Adult Education", IV.4, *Fourth Annual Conference of A.A.A.E., 1964, Proceedings,* quotes a 1964 survey and supports my W.A. findings of 1960-61 (previously reported).
LAWTON, C. R.: Collected data on Italian I and German I students over recent years provides supporting evidence from University of Adelaide.

[3] JONES, H. A., put this idea at the N.I.A.E. 1963 Annual Conference at Exeter, and also recently the Head, Department of Classical and Modern Languages, at "The City Lit.", HAY, M., in "Language Confusion", *Adult Education, N.I.A.E.*, Vol. 39, No. 5, January, 1967, p. 291.

[4] WYKES, O.: "Survey of Foreign Language Teaching" in the Australian Universities (mimeo), Australian Humanities Research Council, November, 1966, Table B, p. 8. Australian students are shown to be dropping languages they have already commenced during secondary school. Persistence rates between Form 1 to Forms 6 or 5 are never higher than 31% in N.S.W. and as low as 5% in Victoria.

[5] But some group contribution can be made even in language class, as suggested by RIVERS, W., "Cultural Understanding Through Languages", p. 15, *op. cit.*
Where tutors are not available, group learning can be applied to languages. The Norwegian discussion group service, Folkets Brevskole, Oslo, has well-produced discussion group courses in Norwegian, English and Esperanto.

[6] LAWRENCE, F., *op. cit.*, p. 51, tried streaming of True beginners from Advanced beginners and found it to be "in vain". I would suggest, nevertheless, that it is worth persisting with the practice until the number of experiments is large enough to allow statistical analysis and the cancelling out of accidental variables.
LAWTON, C. R., has practised streaming at Adelaide since 1962. A.R. 1962 Dept. of Adult Education, The University of Adelaide.

[7] BUROS, O. K., ed.: *Sixth Mental Measurement Year Book,* The Gryphon Press N.J., 1965, pp. 633-637.

[8] *Modern Language Aptitude Test,* JOHN B. CARROL and STANLEY N. SAPON, 1958, The Psychological Corporation N.Y. 17, N.Y.
The quotations and precis are from the *M.L.A.T. Manual,* pp. 3-4; and information from R.A.A.F. School of Languages, Point Cook, Victoria.

[9] Sample 1 consisted of: Italian I, II and III, 49; Malay I, II and III, 24; French I, 14; German I, 21; Russian I, 3. Total = 111 students.
The average attendance in September, 1966, in these classes was 247 students.

AMBASSADOR PRESS—24435